NOAA Central Library

National Oceanic and Atmospheric Administration

Bibliometrics at the NOAA Central Library

Sarah Davis
Jamie Roberts
May 23, 2017

Bibliometrics at the NOAA Central Library

NOAA Central Library
National Oceanic and Atmospheric Administration

Webinar Agenda

- Bibliometrics 101
- Types of metrics
- Process & Tools
- Tracking NOAA Publications
- Other Projects
- Resources

Bibliometrics 101

NOAA Central Library
National Oceanic and Atmospheric Administration

WHAT IT IS…

The scientific and quantitative analysis of academic research; a way of measuring the authorship, publication and use of literature as a proxy for research

BIBLIOMETRICS IN THE LIBRARY…

- Evaluate the output of authors, programs, institutions and countries
- Study areas of scientific research and identify trends
- Assist with library tasks such as evaluating journal titles for collection development
- Aid authors in selecting journals for publication and identifying seminal research in their field

Bibliometrics 101

BASIC PRINCIPLES

- **Bibliometric indicators depend entirely on data quality**
  - In publication collection, proper author and institution disambiguation is crucial to retrieving high quality results.
  - In database selection differences in coverage, search capabilities, associated tools and indexing quality should be given careful consideration.
- **Most bibliometric datasets have skewed distributions**
  - The average is not representative of the dataset and this needs to be accounted for. Providing median as well as mean figures will resolve this to some extent.
- **Citation counts and behaviors vary from field to field and over time**
  - Most metrics cannot be compared between fields.
  - Publications should be at least two years old in order to provide meaningful citation data.
- **Context is Key!**
  - No indicator is perfect and each measures something different so use multiple indicators to give the most complete picture possible.

Bibliometrics 101

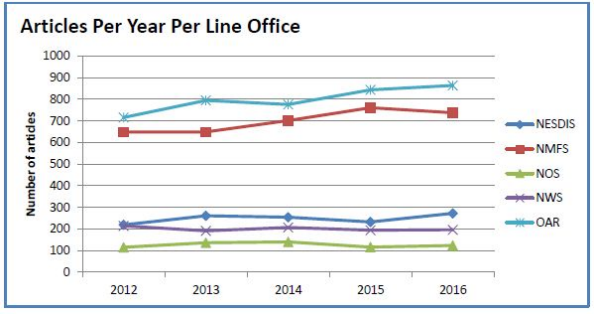
LIMITATIONS AND CONSIDERATIONS

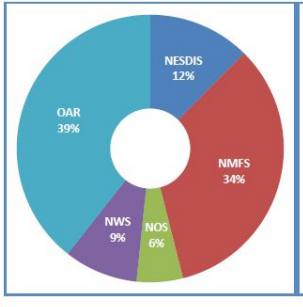
- **Bibliometric indicators are imperfect and can only be used for their specific purpose.**
  - For example: Journal Impact Numbers can only be used to evaluate journals and not the papers in those journals.
- **There's no way to definitively know what indicators actually measure.**
  - For instance, high citation rates don't necessarily correlate to credibility and an article may be highly cited because the research it presents is being questioned.
- **Bibliometric analysis only takes into account publications and not other factors.**
  - These metrics cannot, for example, replace peer review and should not be used as a sole method of evaluation.
- **Many indicators can be gamed.**
  - Journal Impact scores can be manipulated by editors and H-Index can be manipulated through self-citation; take care when relying on certain metrics.
- **Prevent misuse of metrics.**
  - Clearly labeling and captioning visualizations decreases likelihood of misuse or manipulation.
  - Be clear with audience about appropriate uses and limitations of bibliometric indicators.

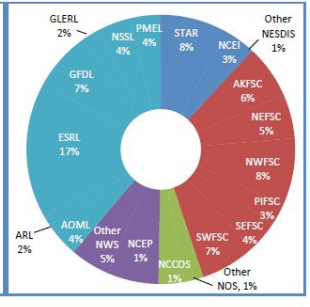
Types of Metrics

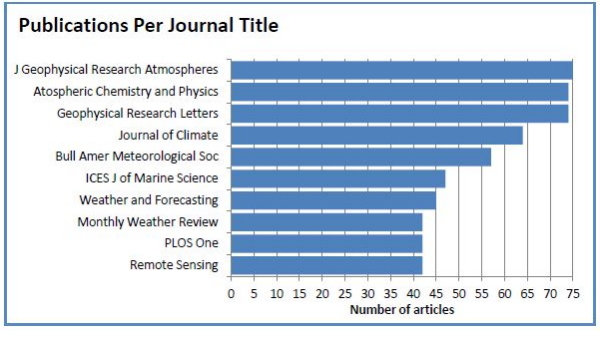
Publication Metrics

- Publication Count
- Publications per:
  - Year or quarter
  - Office or division
  - Journal title
  - Subject area
  - Author
  - Funding Agency
  - Type of publication









Types of Metrics

Citation Metrics

- Citation Rate
- Citations Per Year
- Mean Citation Rate
- Median Citation Rate
- H-Index

NOTE: Articles require at least 2 years to accumulate enough citations for article-level bibliometric indicators to be reliable.

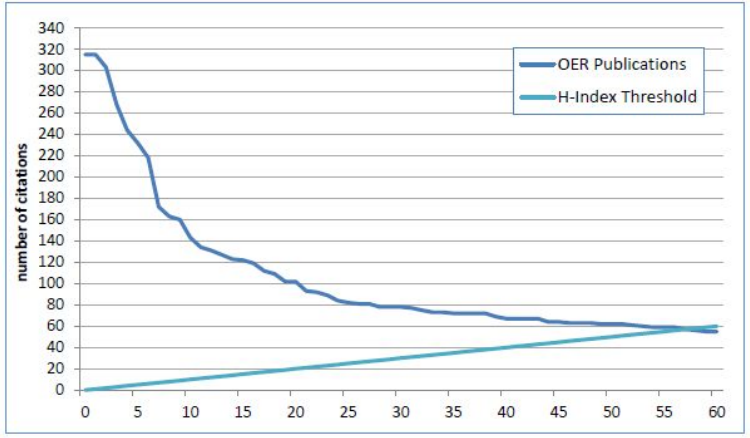
Citation metrics from WoS

Results found:	1438
Sum of the Times Cited [?] :	83319
Sum of Times Cited without self-citations [?] :	79578
Citing Articles [?] :	54487
Citing Articles without self-citations [?] :	53459
Average Citations per Item [?] :	57.94
h-index [?] :	118

H-Index

"A scientist has index h if h of his/her N_p papers have at least h citations each, and the other (N_p-h) papers have no more than h citations each."

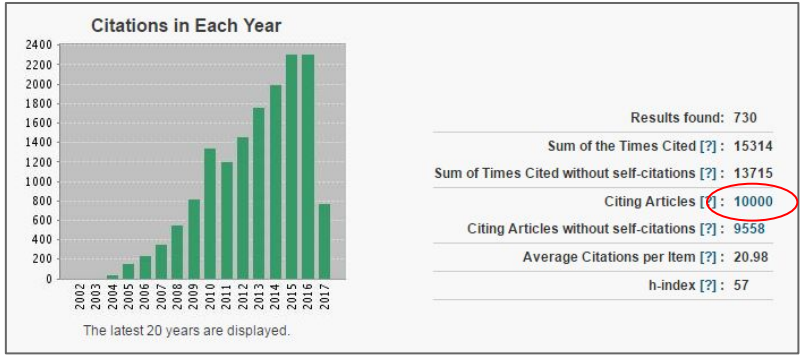
Hirsch, J.E. (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA, 102(46), 16569-16572. doi:10.1073/pnas.0507655102



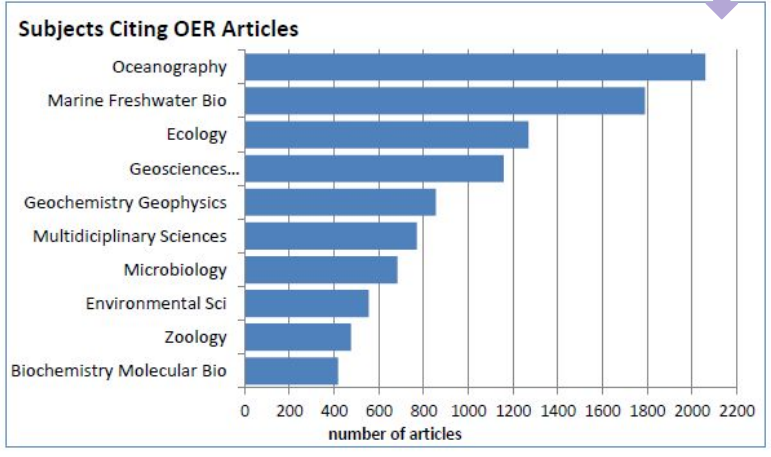
Types of Metrics

Analysis of Citing Articles

- Simply publication metrics calculated for the articles that cited our articles.
- Provides a picture of how NOAA research is being used by the greater scientific community.
- Metrics include:
  - Citing articles per subject and journal
  - Institutions and funding agencies
  - International publications



Clicking on the Citing Article in the WoS Citation Report creates a new set of articles that can be analyzed.



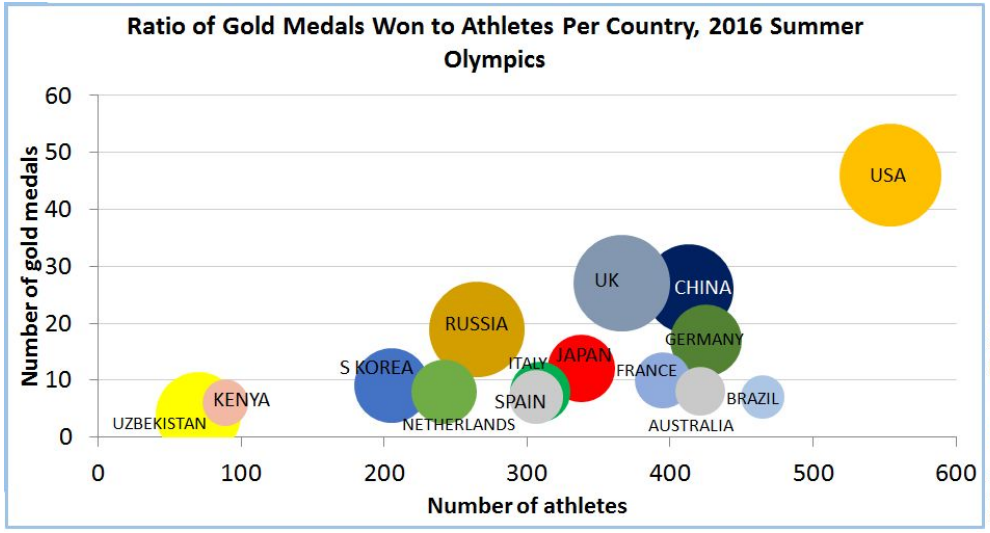
Types of Metrics

Percentile Ranks

Calculating percentile ranks based on distribution of citation counts to all papers in a given set, normalizes citations and thus can give better comparisons across disciplines

- Can be used for evaluation, comparison, and ranking.
- Tools such as Essential Science Indicators can help to calculate percentile ranks.

- Percentage of articles in top 1% or 10% of all articles in a given field help illustrate high quality of research.



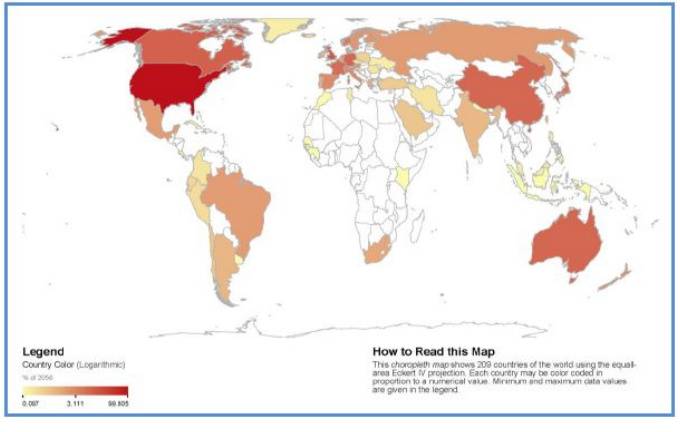
Types of Metrics

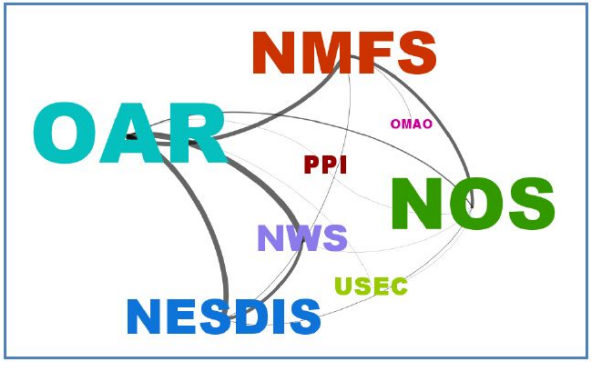
Collaboration Rates and Networks

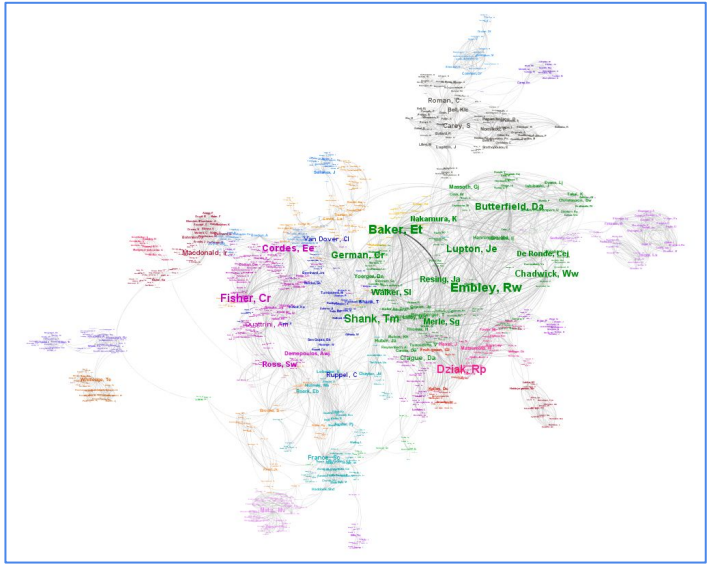
Shows collaborations between authors, institutions, countries, etc.

COLLABORATION

Type of Collaboration	Rate
Intramural collaboration at the line office level	5.76 %
Intramural collaboration at the research unit level	12.49 %
Extramural collaboration at the institutional level	94.02 %
Extramural collaboration at the international level	44.02 %



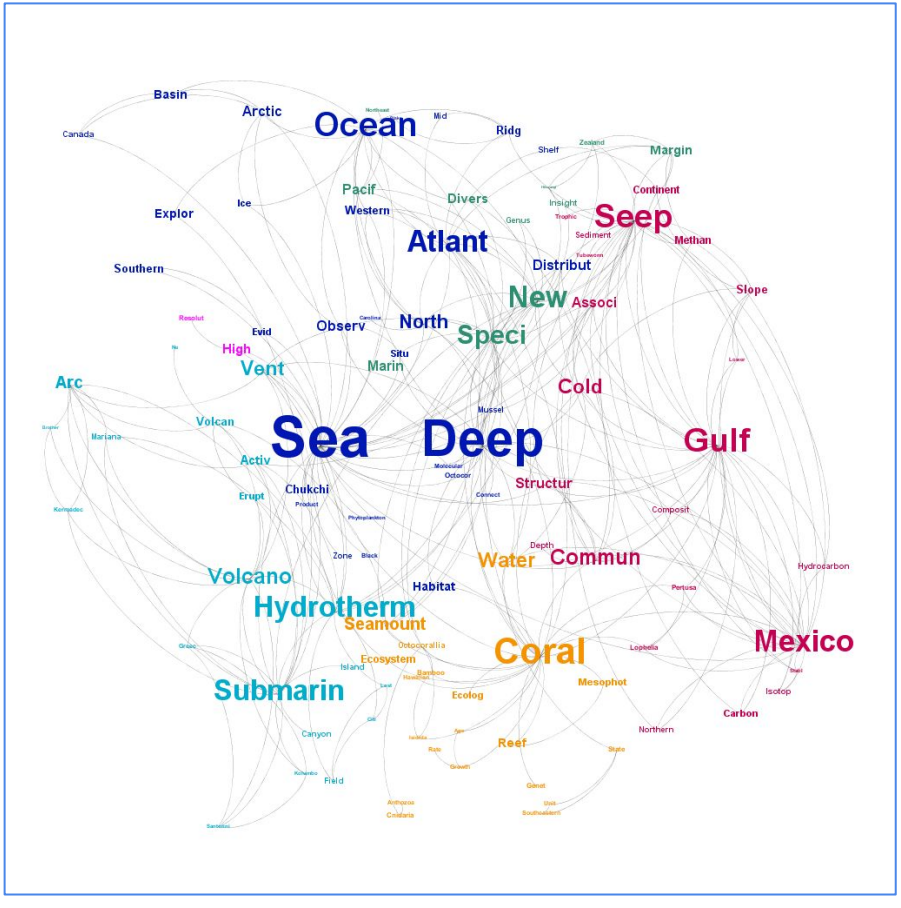




Types of Metrics

Semantic Networks

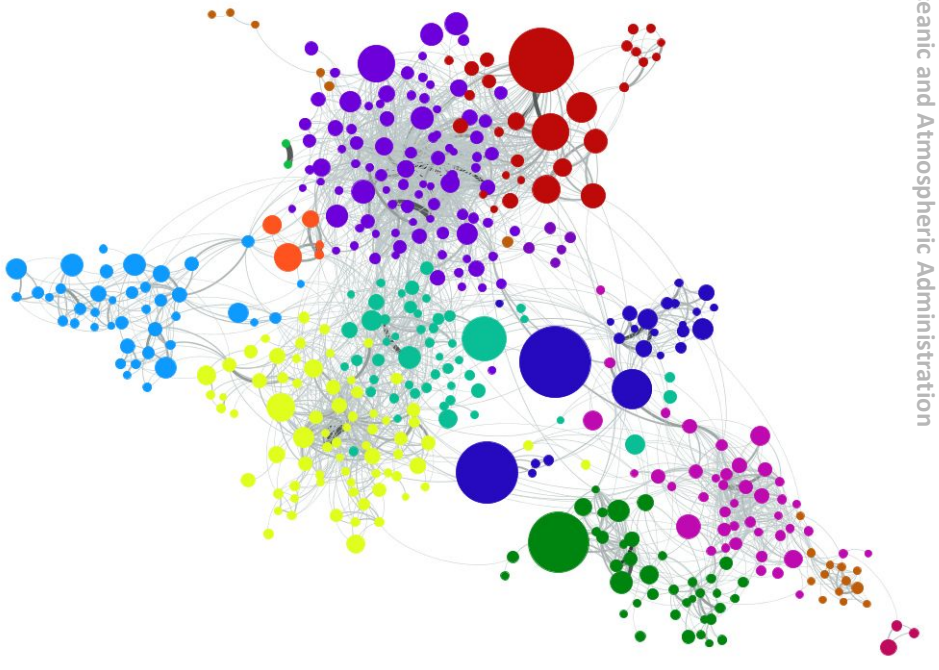
- Show the most commonly co-occurring words in the titles, abstracts, keywords, etc. of articles in a given dataset.
- Useful in illustrating common themes or subjects of research and how those interrelate.



Types of Metrics

Citation Networks

- Networks illustrating relationships between publications cited by articles in a set.
- Three types of citation networks:
  - Direct - edges connect articles that cite each other.
  - Bibliographic coupling - edges connect articles that cite the same publications.
  - Co-citation - edges connect articles cited by the same publications.
- Nodes always represent publications but edges represent different relationships in each case.



TYPES OF METRICS

Journal Level Metrics

- Seek to illustrate relative importance of a journal within its field.
- Basic indicators include:
  - **Journal Impact Factor (JIF)**
  - **Eigenfactor**
  - Scimago Journal & Country Rank (SJR)
  - Cited Half-Life
  - CiteScore (Scopus)
- Not appropriate for evaluating articles or researchers.

Formula for JIF

J.I.F. 2012= Citations Received in 2012 to Items Published in 2010 or 2011

Citable Items in 2010 and 2011

InCites Journal Citation Report

		Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
☐	9	NATURE	627,846	38.138	1.44256

Cited Half-Life Data

Cited Year	2015	2014	2013	2012	2011
#Cites from 2015	8,538	31,056	34,618	42,643	36,247
Cumulative %	1.36	6.31	11.82	18.61	24.39

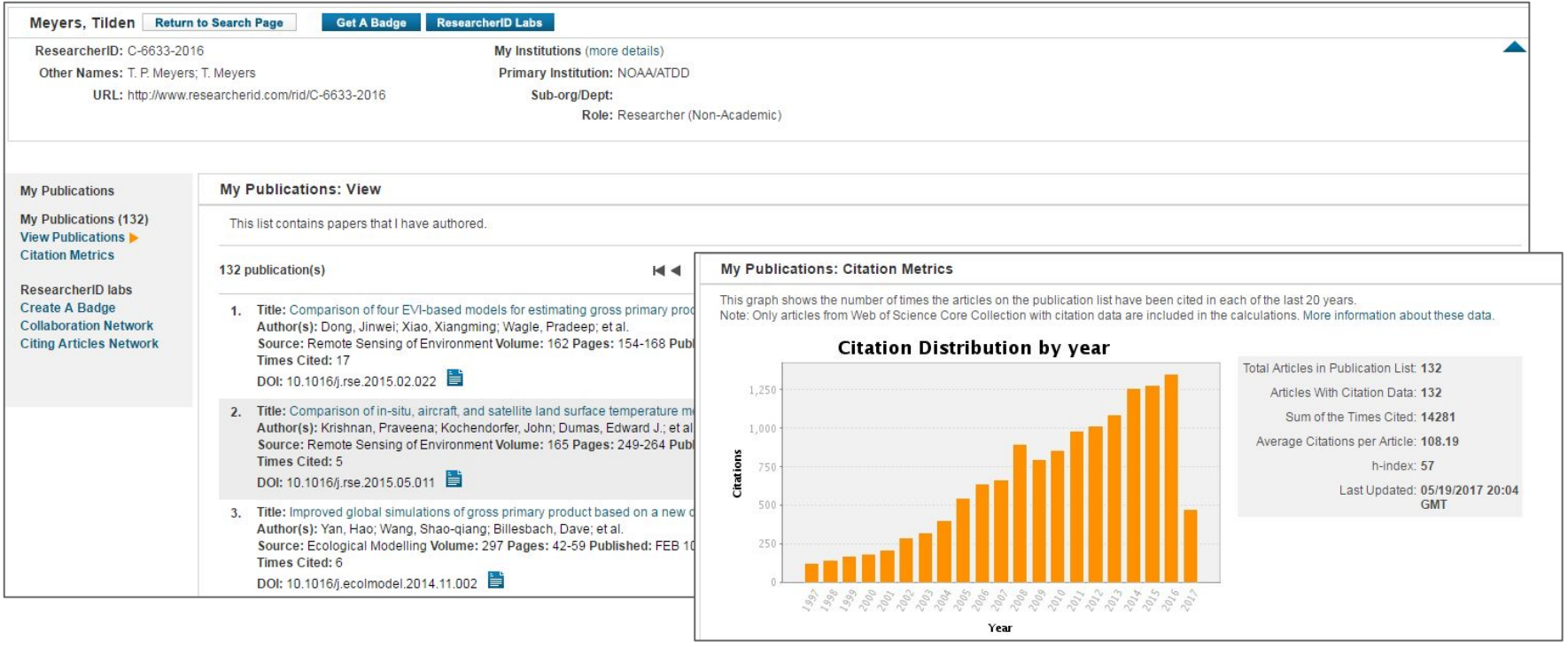
Eigenfactor Journal Ranking

NOAA Central Library
National Oceanic and Atmospheric Administration

Types of Metrics

Author Metrics In ResearcherID

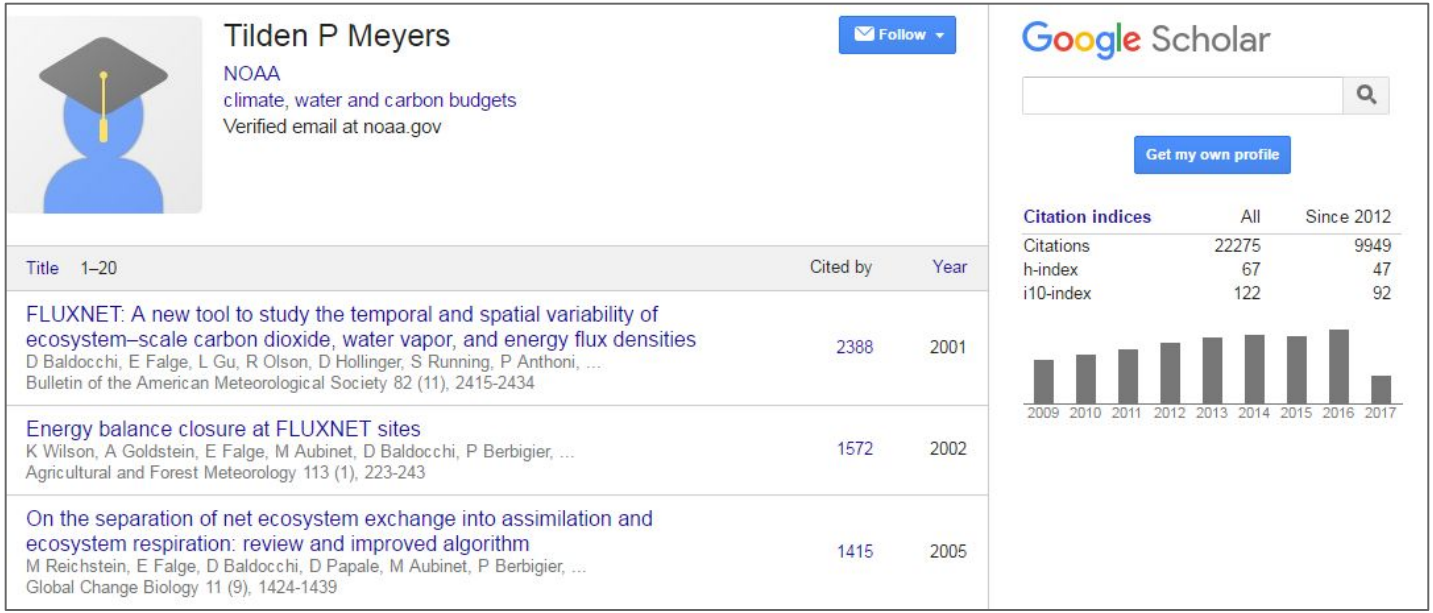
Authors who have ResearcherIDs have access to basic citation metrics as well as ResearcherID labs and can use these tools to calculate their own metrics.



Types of Metrics

Author Metrics In Google Scholar

Google Scholar is difficult to use when analyzing the output of an office or program but does provide some useful, albeit noisy, metrics for author level analysis and can provide insight into wider use of research in both peer-reviewed and non-peer-reviewed sources.



Tilden P Meyers

Follow

NOAA

climate, water and carbon budgets

Verified email at noaa.gov

Title 1–20	Cited by	Year
FLUXNET: A new tool to study the temporal and spatial variability of ecosystem–scale carbon dioxide, water vapor, and energy flux densities D Baldocchi, E Falge, L Gu, R Olson, D Hollinger, S Running, P Anthoni, ... Bulletin of the American Meteorological Society 82 (11), 2415-2434	2388	2001
Energy balance closure at FLUXNET sites K Wilson, A Goldstein, E Falge, M Aubinet, D Baldocchi, P Berbigier, ... Agricultural and Forest Meteorology 113 (1), 223-243	1572	2002
On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm M Reichstein, E Falge, D Baldocchi, D Papale, M Aubinet, P Berbigier, ... Global Change Biology 11 (9), 1424-1439	1415	2005

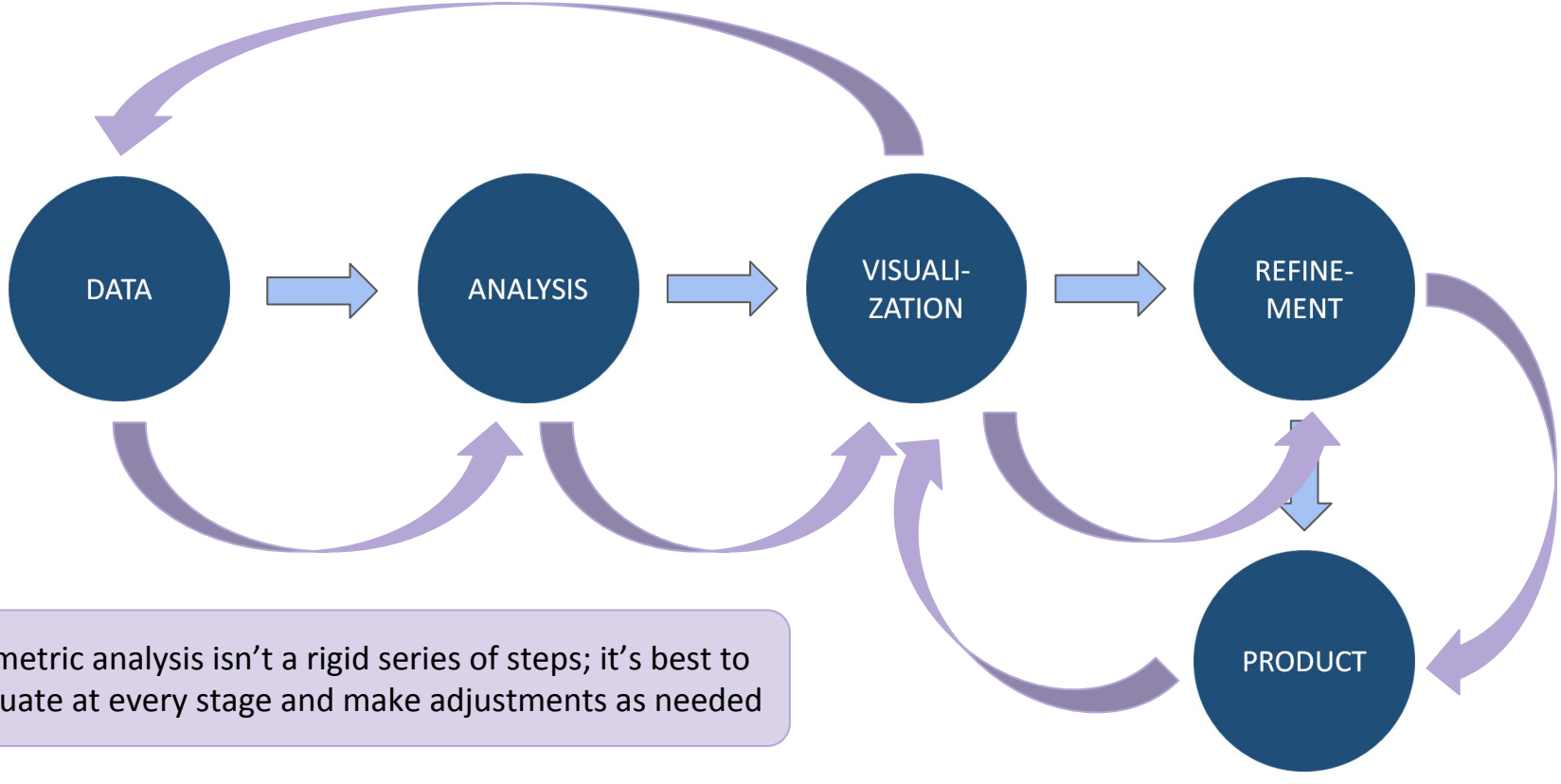
Google Scholar

Get my own profile

Citation indices	All	Since 2012
Citations	22275	9949
h-index	67	47
i10-index	122	92

2009 2010 2011 2012 2013 2014 2015 2016 2017

The Process



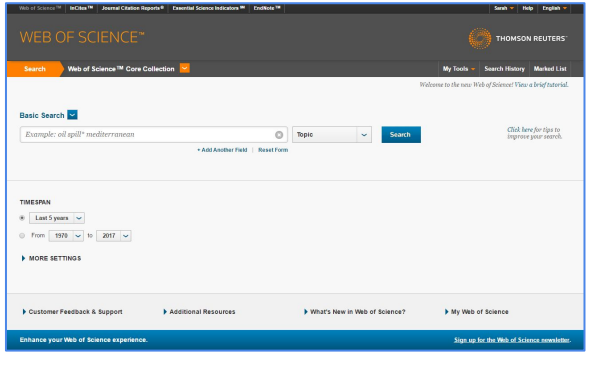
Bibliometric analysis isn't a rigid series of steps; it's best to reevaluate at every stage and make adjustments as needed

Tools For Bibliometrics

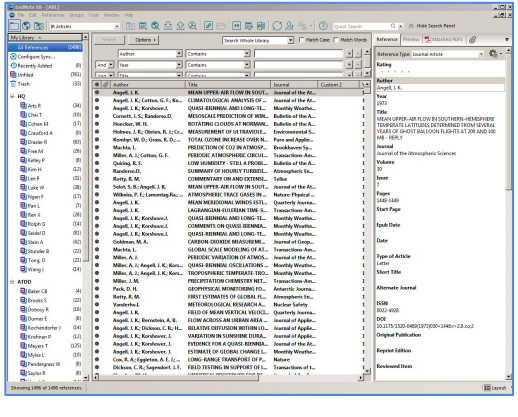
Our Basic Bibliometric Toolbox

For Data Collection and Management

Web of Science

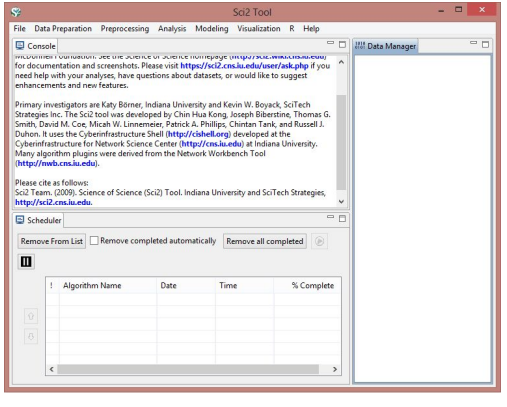


EndNote

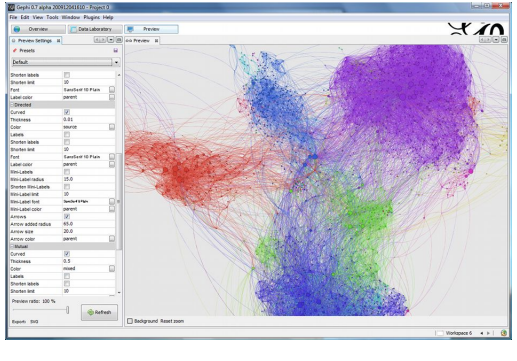


For Data Analysis and Visualization

Science of Science Tool (Sci2)



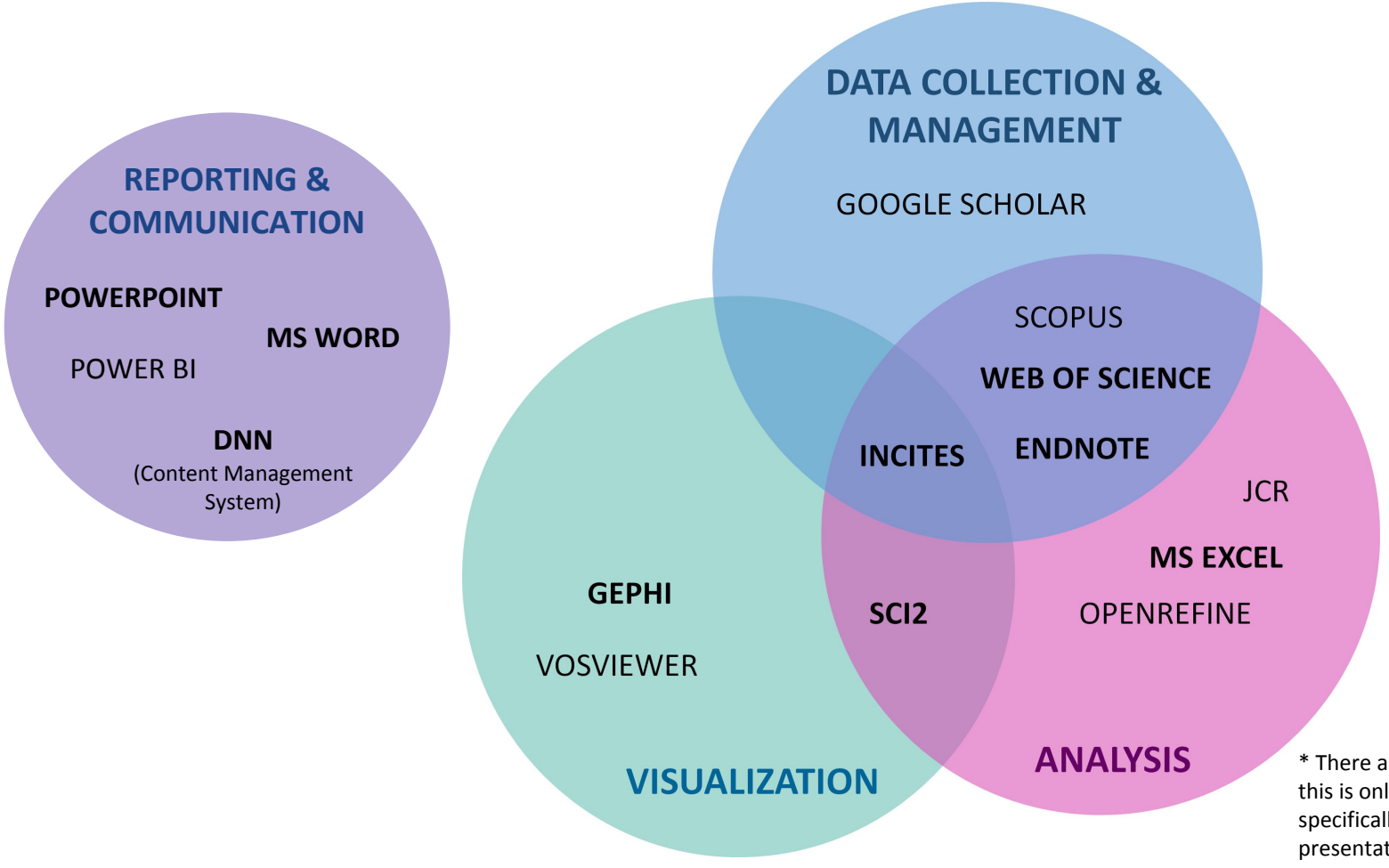
Gephi



NOAA Central Library

National Oceanic and Atmospheric Administration

Tools For Bibliometrics



* There are many tools available, this is only a selection. Tools specifically discussed in this presentation are in bold text.

NOAA-Wide Publications Tracking

The NOAA-Wide Publications Tracking seeks to report the number of peer-reviewed articles produced by NOAA authors in a consistent and reproducible manner.

- Started in 2012, and added funded articles in 2016
- More than 10,000 NOAA-authored articles collected to date
- Metrics produced are reported quarterly and are used in the balanced scorecard which monitors NOAA's performance.

Our Team

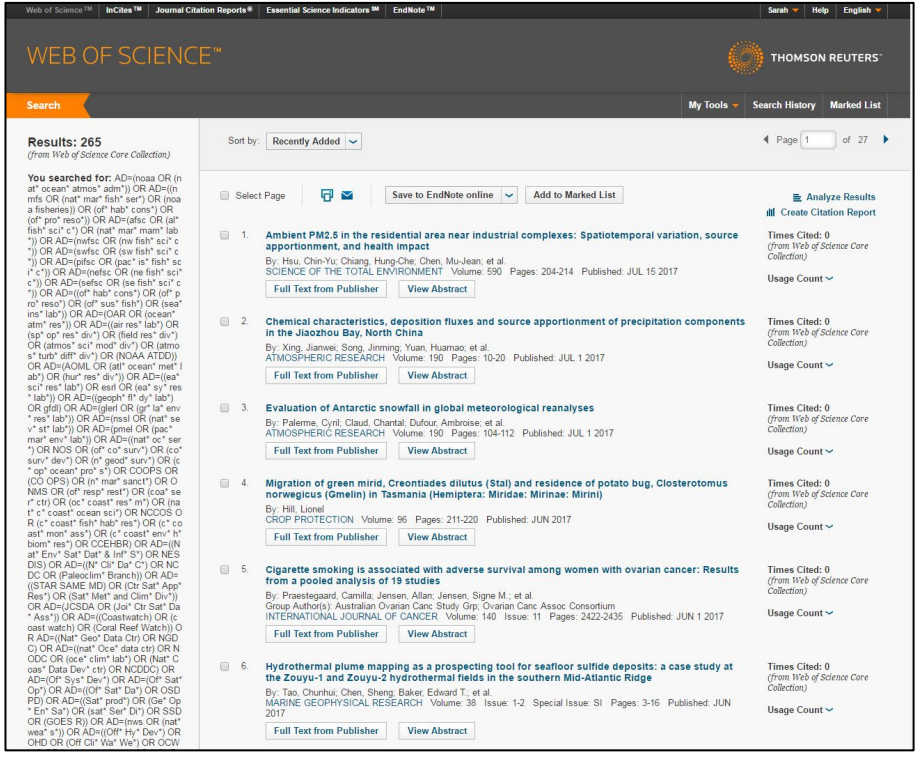
Sarah Davis
Jamie Roberts
Jan Thomas (College Park)
Ashley Jefferson (Miami)

NOAA Central Library
National Oceanic and Atmospheric Administration

NOAA-Wide Publications Tracking

DATA COLLECTION

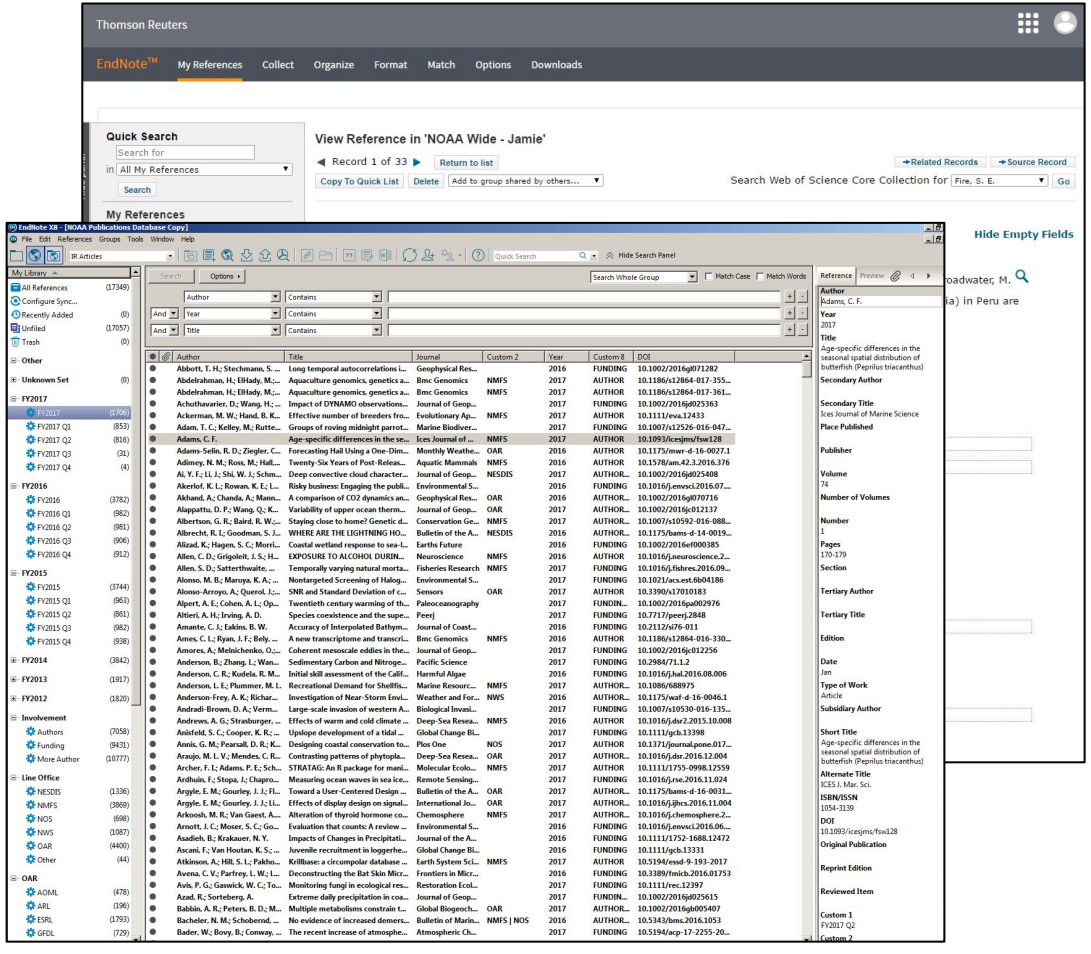
- Weekly report run in Web of Science using saved search string.
- New results saved to EndNote online for review by the bibliometrics team.



NOAA-Wide Publications Tracking

DATA REVIEW

- Bibliometrics team reviews records in EndNote online and adds project specific metadata.
- Edited records are saved to EndNote desktop and given a final quality control review.
- Periodically, further review in an effort to “clean up” data.



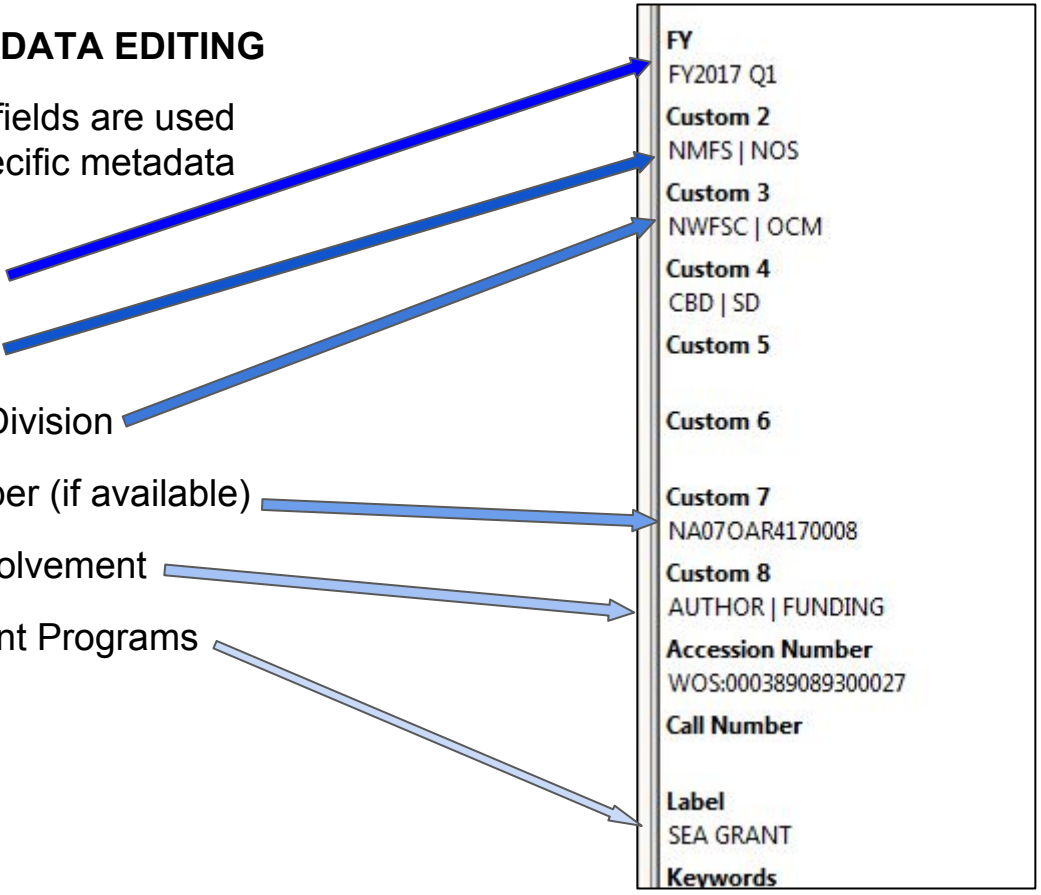
NOAA Central Library

National Oceanic and Atmospheric Administration

NOAA-Wide Publications Tracking

DATA REVIEW: METADATA EDITING

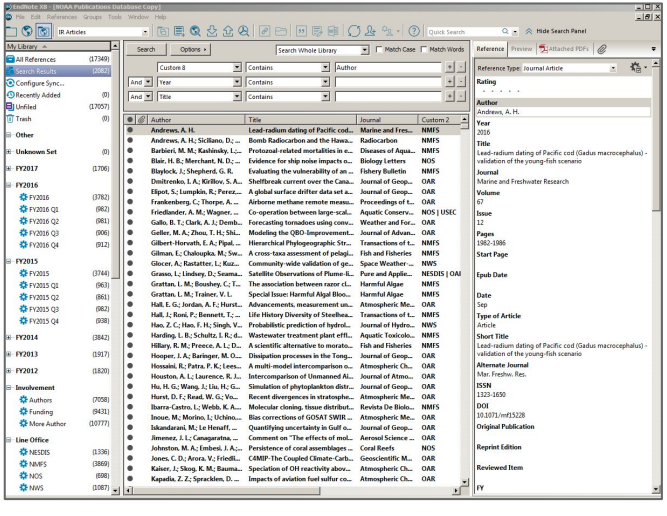
- EndNote custom fields are used to add project-specific metadata including
  - Fiscal Year
  - Line Office
  - Lab/Office/Division
  - Grant Number (if available)
  - NOAA's Involvement
  - Certain Grant Programs

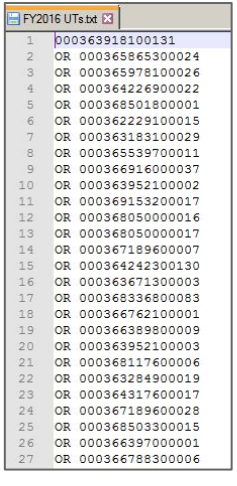


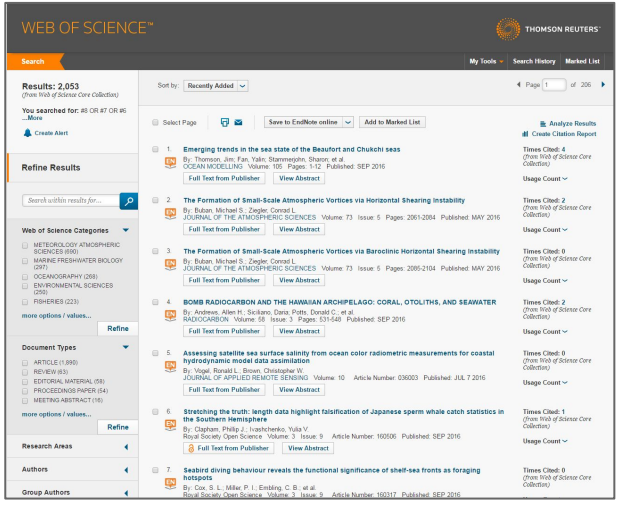
NOAA-Wide Publications Tracking

DATA ANALYSIS & VISUALIZATION

Most analyses start with creating a dataset within our EndNote database, exporting a list of accession numbers, and then running a search for those numbers in Web of Science. From there the **Analyze Results** and **Create Citation Report** functions offer a number of tools for creating metrics.







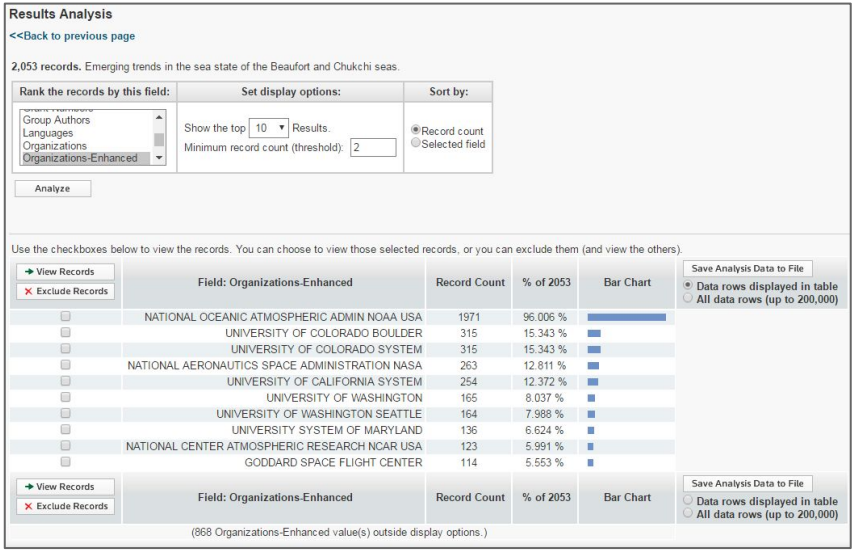
NOAA Central Library

National Oceanic and Atmospheric Administration

NOAA-Wide Publications Tracking

DATA ANALYSIS & VISUALIZATION

Using the analysis tools in Web of Science we can create basic publication and citation metrics which can easily be visualized using Excel.



Results Analysis

<<Back to previous page

2,053 records. Emerging trends in the sea state of the Beaufort and Chukchi seas.

Rank the records by this field: Group Authors, Languages, Organizations, Organizations-Enhanced

Set display options: Show the top 10 Results. Minimum record count (threshold): 2

Sort by: Record count / Selected field

Analyze

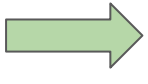
Use the checkboxes below to view the records. You can choose to view those selected records, or you can exclude them (and view the others).

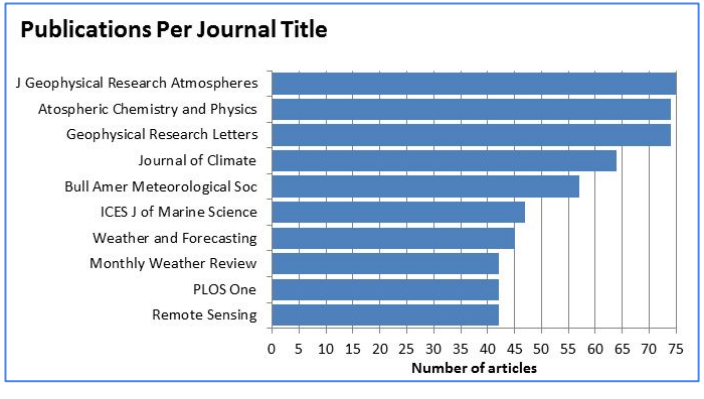
View Records / Exclude Records

Field: Organizations-Enhanced	Record Count	% of 2053	Bar Chart
NATIONAL OCEANIC ATMOSPHERIC ADMIN NOAA USA	1971	96.006 %	
UNIVERSITY OF COLORADO BOULDER	315	15.343 %	
UNIVERSITY OF COLORADO SYSTEM	315	15.343 %	
NATIONAL AERONAUTICS SPACE ADMINISTRATION NASA	263	12.811 %	
UNIVERSITY OF CALIFORNIA SYSTEM	254	12.372 %	
UNIVERSITY OF WASHINGTON	165	8.037 %	
UNIVERSITY OF WASHINGTON SEATTLE	164	7.988 %	
UNIVERSITY SYSTEM OF MARYLAND	136	6.624 %	
NATIONAL CENTER ATMOSPHERIC RESEARCH NCAR USA	123	5.991 %	
GODDARD SPACE FLIGHT CENTER	114	5.553 %	

Save Analysis Data to File: Data rows displayed in table / All data rows (up to 200,000)

(868 Organizations-Enhanced value(s) outside display options.)

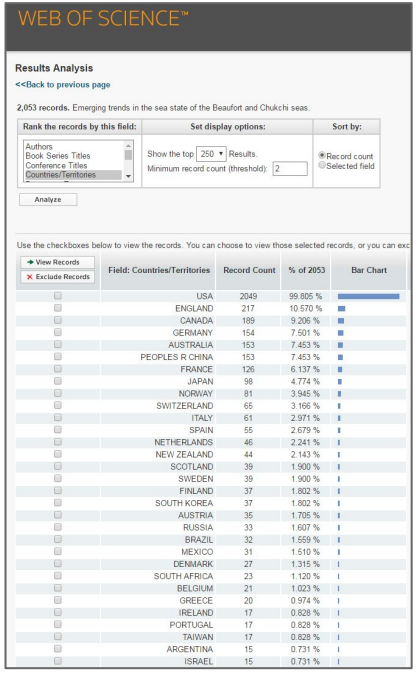


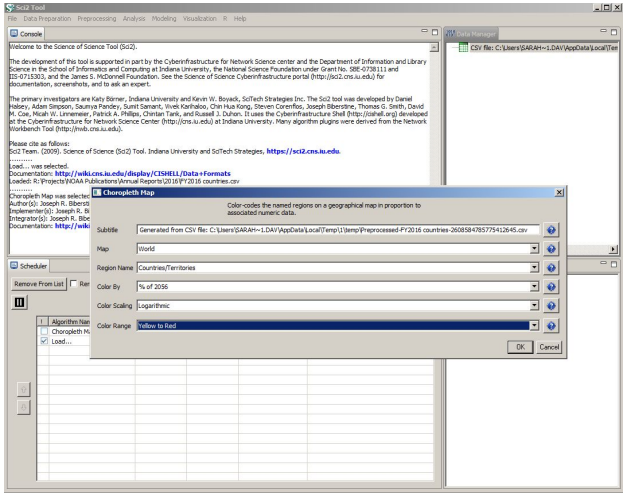


NOAA-Wide Publications Tracking

DATA ANALYSIS & VISUALIZATION

Geographic maps are created by analyzing results in Web of Science by territory, saving that data as a .csv file which can then be loaded into the Sci2 tool which has the ability to analyze that data and create a geographic visualization.





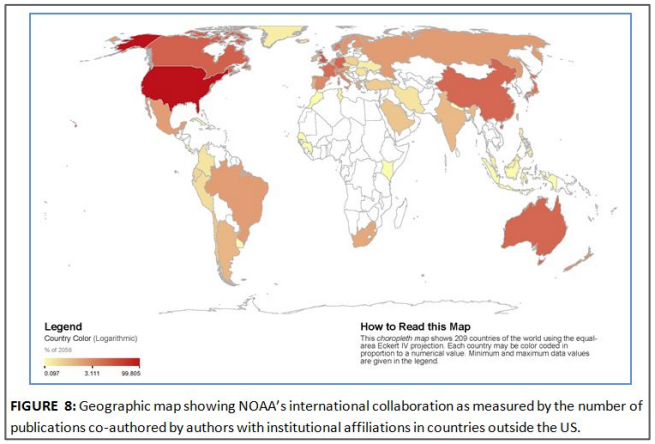


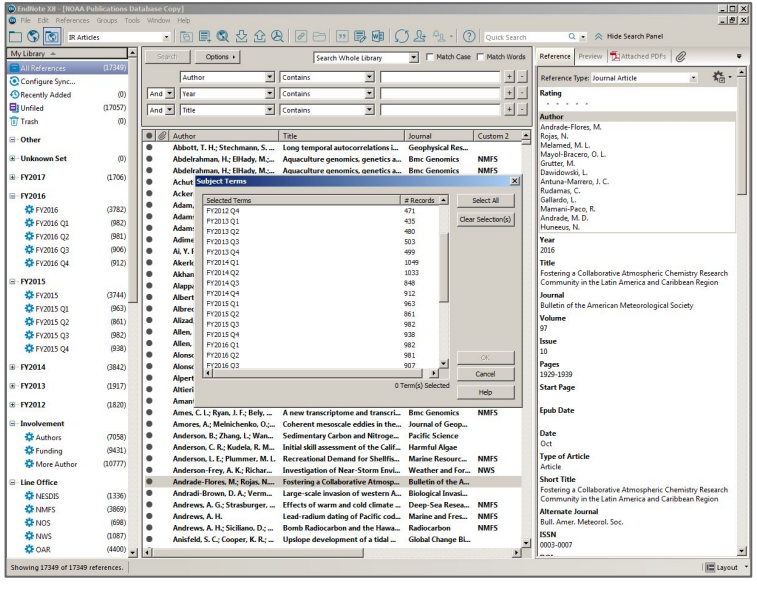
FIGURE 8: Geographic map showing NOAA's international collaboration as measured by the number of publications co-authored by authors with institutional affiliations in countries outside the US.

NOAA-Wide Publications Tracking

NOAA Central Library
National Oceanic and Atmospheric Administration

DATA ANALYSIS & VISUALIZATION

The metadata added during data collection and review such as fiscal year and line office can be analyzed using the tools in EndNote desktop and visualized using Excel.



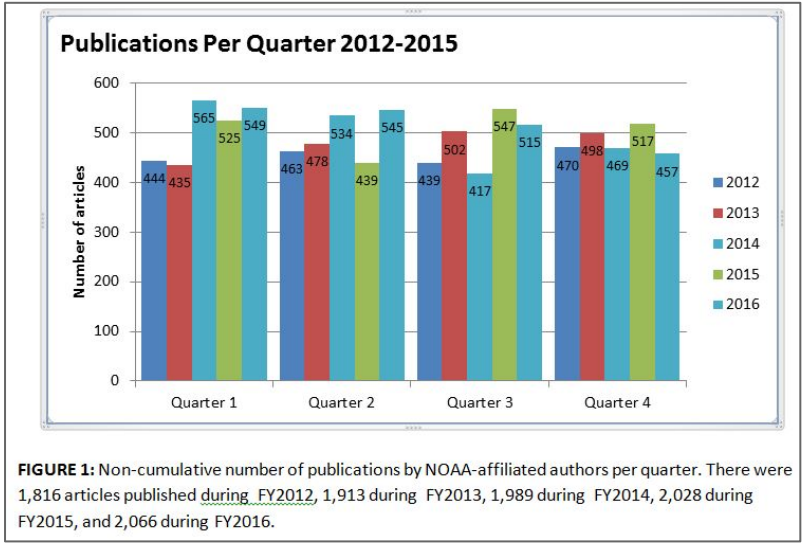
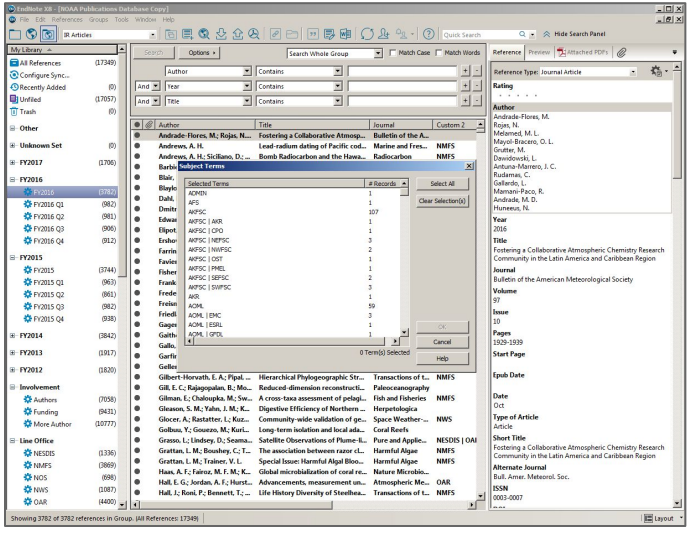


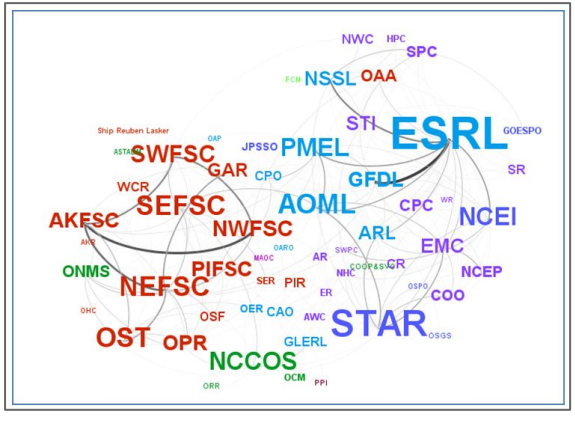
FIGURE 1: Non-cumulative number of publications by NOAA-affiliated authors per quarter. There were 1,816 articles published during FY2012, 1,913 during FY2013, 1,989 during FY2014, 2,028 during FY2015, and 2,066 during FY2016.

NOAA-Wide Publications Tracking

DATA ANALYSIS & VISUALIZATION

The line office and lab metadata added during data collection can also used to create collaboration network maps. This is done by exporting that data into Excel, loading it into Sci2 for analysis and then visualizing that analysis in Gephi.

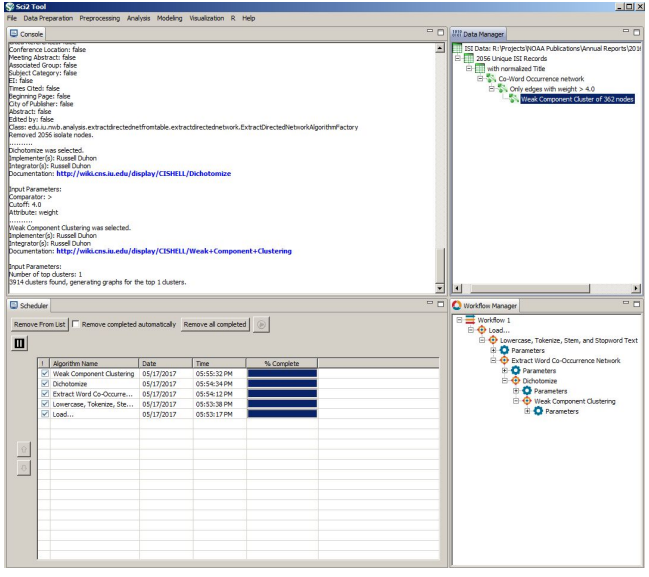


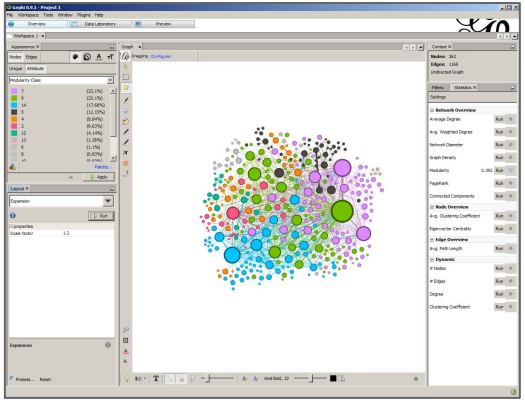


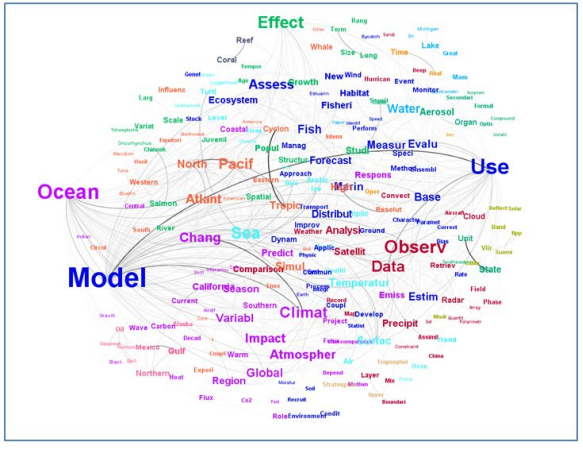
NOAA-Wide Publications Tracking

DATA ANALYSIS & VISUALIZATION

Web of Science search results can be saved as an .isi file which can then be loaded into the Sci2 tool for network analysis. Once analyzed in Sci2, networks are visualized in Gephi and saved as .png files. Visualizations are also saved as .pdf files for potential use in other formats. WoS search results can also be saved as .csv files for



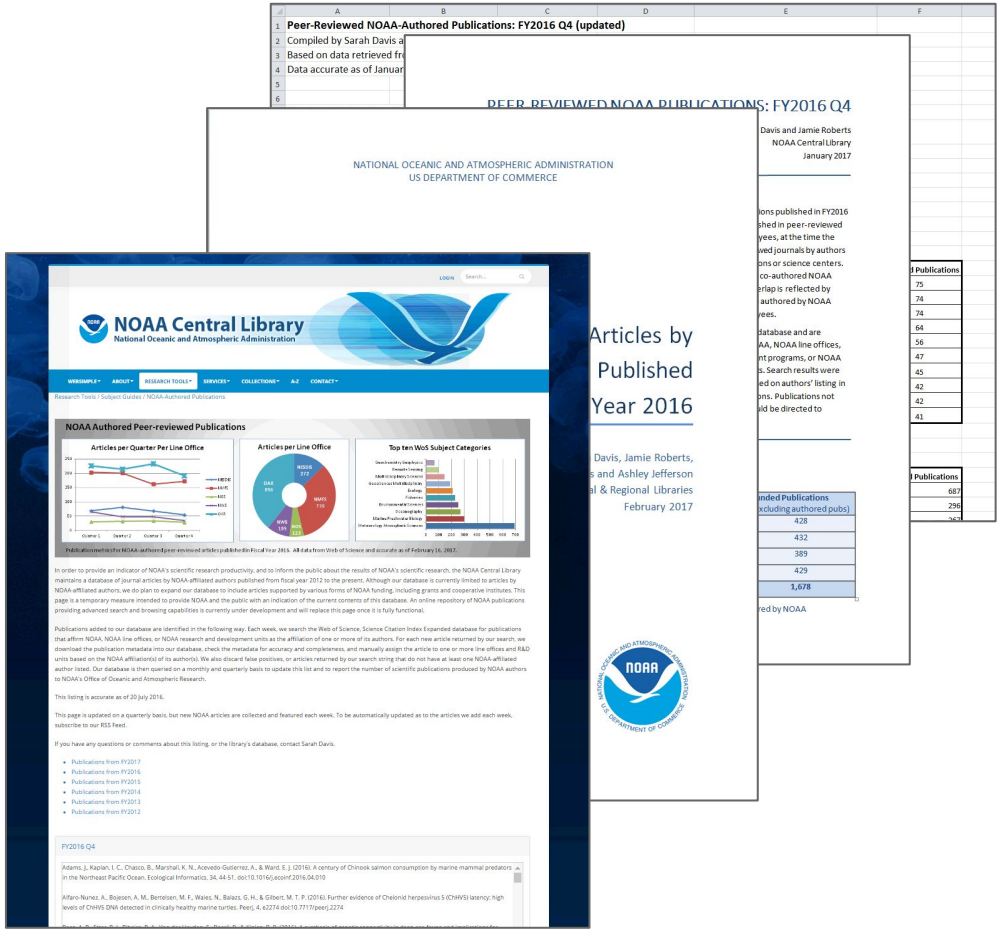




NOAA-Wide Publications Tracking

Project Deliverables

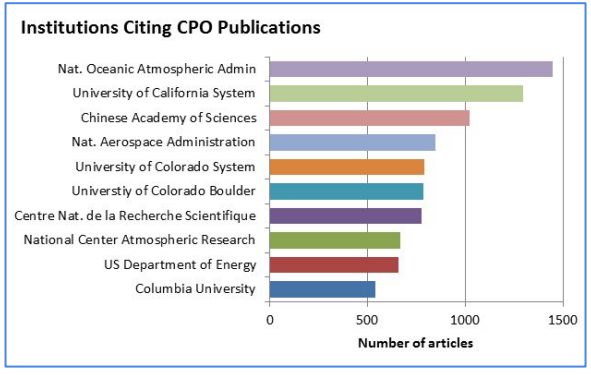
- Quarterly Excel report with list of articles by line office and basic metrics
- Executive summary of quarterly publications report
- Annual publication analysis report
- Online bibliography of article organized by quarter with dashboard and links to annual reports

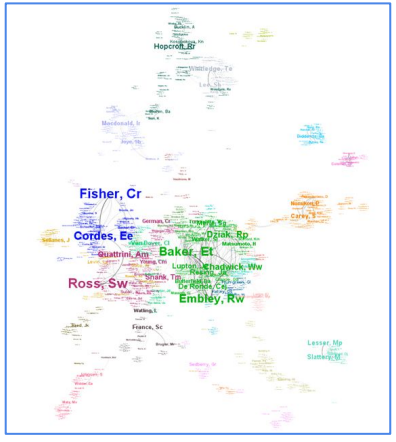


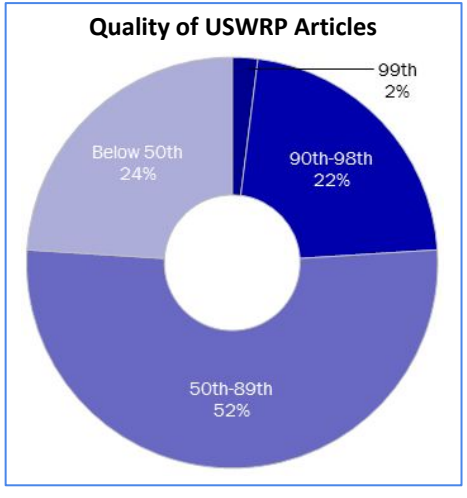
Other Projects

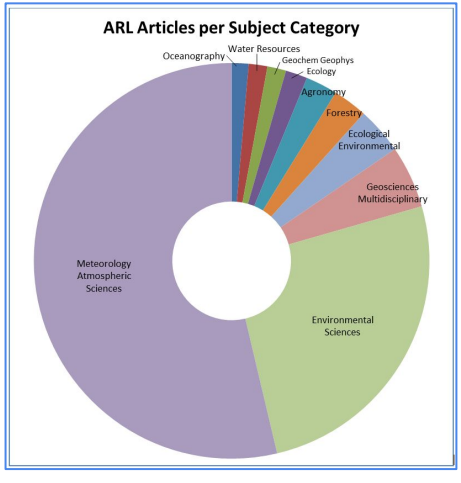
Both ongoing and one-time projects

- Most projects are requested by a program or grant office.
- Typically, data is parsed from our existing database *or* client provides a list of publications.
- We prefer datasets of over 100 articles; larger datasets allow for more meaningful analysis.
- A generic checklist provides a starting point for new projects and each project is documented and all incidental materials and data are saved in a common location.









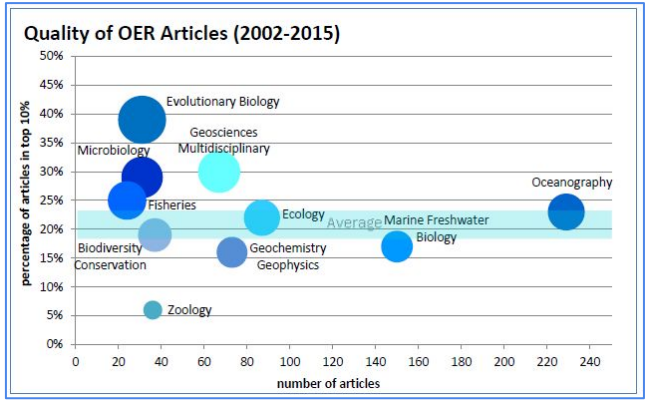
Other Projects

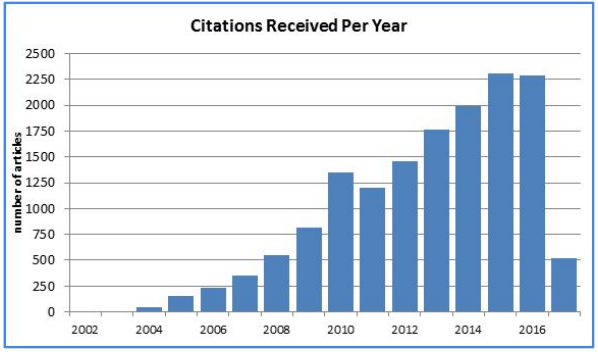
Office of Ocean Exploration and Research (OER) Citation Report

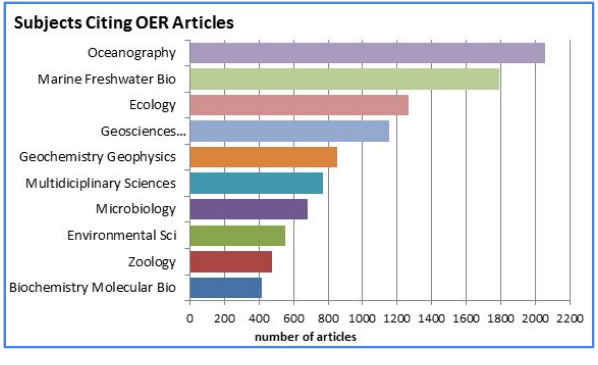
- Produced quarterly and analyzing peer-reviewed articles 2002-present
- Dataset derived from a Web of Science search as well as a list of publications provided by OER

SUMMARY METRICS

Bibliometric Indicator	Value
Number of Publications (p)	730
Total Number of Citations Received (c)	15039
Average Number of Citations per Paper (c/p)	20.6
H- Index	57
Percentage of Publications in the Top 10% for Citation Counts	≈21.5%







NOAA Central Library
National Oceanic and Atmospheric Administration

Promoting Our Program

Webpage

Single page with pop-ups that:

- Explain bibliometrics;
- Showcase our products and services;
- Connect users to tools available to do their own analyses.

One-Sheet

Simplified content that reflects what we feature on the web page.

Brown Bag & Conference Presentations

Facebook & NOAA-wide emails

Resources

Open Source Tools We Use

Science of Science (Sci2) Tool a toolset designed for the study of science that supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets.

Download the tool at: https://sci2.cns.iu.edu/user/index.php (you will need to register for an account)

The Sci2 Manual provides details about the tool, helpful tutorials and sample data sets for learning purposes.

The book Visual Insights: A Practical Guide to Making Sense of Data by Katy Börner and Ted Polley (2014, Cambridge, MA: The MIT Press.) offers an introduction to visualization design through both theory and practical application.

Gephi an open-source network analysis and visualization tool.

Download the tool at: https://gephi.org/

The Gephi User Guide offers an introduction to networks and network analysis, tutorials and sample data sets.

A Gephi Facebook group provides a link to experienced users and discussions as well as examples of successful visualizations created using the tool.

Resources

Other Open Source Tools Available

VOSViewer http://www.vosviewer.com/

Pajek Wiki http://pajek.imfm.si

CiteSpace http://cluster.cis.drexel.edu/~cchen/citespace

VOSviewer www.vosviewer.com

NodeXL http://nodexl.codeplex.com

NetDraw http://sites.google.com/site/netdrawsoftware

Cytoscape www.cytoscape.org

Google Charts https://developers.google.com/chart/

OpenRefine http://openrefine.org/

Sarah Davis
sarah.davis@noaa.gov

Jamie Roberts
jamie.roberts@noaa.gov